

A Supervised Segmentation Network for Hyperspectral Image Classification

Hao Sun^{ID}, Xiangtao Zheng^{ID}, Member, IEEE, and Xiaoqiang Lu^{ID}, Senior Member, IEEE

Abstract—Recently, deep learning has drawn broad attention in the *hyperspectral image* (HSI) classification task. Many works have focused on elaborately designing various spectral-spatial networks, where *convolutional neural network* (CNN) is one of the most popular structures. To explore the spatial information for HSI classification, pixels with its adjacent pixels are usually directly cropped from hyperspectral data to form HSI cubes in CNN-based methods. However, the spatial land-cover distributions of cropped HSI cubes are usually complicated. The land-cover label of a cropped HSI cube cannot simply be determined by its center pixel. In addition, the spatial land-cover distribution of a cropped HSI cube is fixed and has less diversity. For CNN-based methods, training with cropped HSI cubes will result in poor generalization to the changes of spatial land-cover distributions. In this paper, an end-to-end *fully convolutional segmentation network* (FCSN) is proposed to simultaneously identify land-cover labels of all pixels in a HSI cube. First, several experiments are conducted to demonstrate that recent CNN-based methods show the weak generalization capabilities. Second, a fine label style is proposed to label all pixels of HSI cubes to provide detailed spatial land-cover distributions of HSI cubes. Third, a HSI cube generation method is proposed to generate plentiful HSI cubes with fine labels to improve the diversity of spatial land-cover distributions. Finally, a FCSN is proposed to explore spectral-spatial features from finely labeled HSI cubes for HSI classification. Experimental results show that FCSN has the superior generalization capability to the changes of spatial land-cover distributions.

Index Terms—Hyperspectral image (HSI) classification, fully convolutional segmentation network (FCSN), generalization.

I. INTRODUCTION

HYPERSPECTRAL imaging instruments can capture rich spectral signatures and intricate spatial information of

Manuscript received May 7, 2020; revised October 3, 2020; accepted January 22, 2021. Date of publication February 4, 2021; date of current version February 12, 2021. This work was supported in part by the National Science Fund for Distinguished Young Scholars under Grant 61925112, in part by the National Natural Science Foundation of China under Grant 61806193 and Grant 61772510, in part by the Innovation Capability Support Program of Shaanxi under Grant 2020KJXX-091 and Grant 2020TD-015, in part by the Natural Science Basic Research Program of Shaanxi under Grant 2019JC-340 and 2019JC-23, and in part by CAAI-Huawei MindSpore Open Fund. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Jocelyn Chanussot. (*Corresponding author: Xiangtao Zheng*.)

Hao Sun is with the Key Laboratory of Spectral Imaging Technology CAS, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, China, and also with the University of Chinese Academy of Sciences, Beijing 100049, China.

Xiangtao Zheng and Xiaoqiang Lu are with the Key Laboratory of Spectral Imaging Technology CAS, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, China (e-mail: xiangtaoz@gmail.com).

Digital Object Identifier 10.1109/TIP.2021.3055613

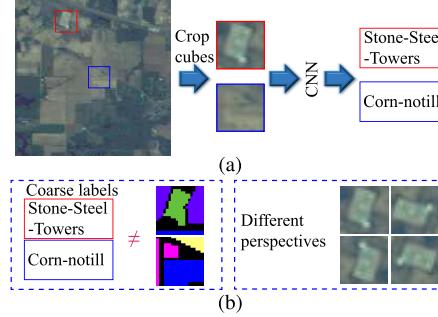


Fig. 1. (a) In CNN-based methods, HSI cubes are directly cropped from the original images and then fed to a CNN to identify the land-cover labels of the center pixels. (b) However, in complicated observed scenes, the land-cover label of a HSI cube cannot be simply determined by its center pixel for training. In addition, the same scene may have different spatial land-cover distributions under different viewing perspectives. The coarse label of a cropped HSI cube cannot provide spatial land-cover distributions for HSI classification.

observed scenes to generate 3D *hyperspectral images* (HSIs). Plentiful spectral signatures and spatial information of HSIs offer great potentials for land-cover mapping [1], [2] and detection [3], [4]. The HSI classification task aims to identify the land-cover category of each hyperspectral pixel in HSIs. However, HSIs are often disturbed by atmospheric disturbance, incident illumination and instrument effects [5], [6]. It is difficult to identify the land-cover categories of hyperspectral pixels.

Recently, many works [7]–[9] attempt to employ the deep learning for HSI classification. According to whether adjacent pixels are explored in deep networks, recent deep learning-based methods are simply summarized into two categories: spectral-based methods and spectral-spatial-based methods. Spectral-based methods focus on exploring spectral signatures of a single pixel for HSI classification [10]. However, the spatial information of hyperspectral data is often ignored by spectral-based methods [11]. Spectral-spatial-based methods usually exploit adjacent hyperspectral pixels for HSI classification [12], [13]. In HSIs, each pixel may share the same land-cover category with its adjacent pixels [14], [15]. These adjacent pixels can provide critical spatial information to classify the hyperspectral pixel. In spectral-spatial-based methods, the most popular deep network is the *convolutional neural network* (CNN) [14], [16]. First, as shown in Fig. 1(a), to explore the spatial information of a hyperspectral pixel, its adjacent pixels within a square area are cropped from original images to form the HSI cube. Then, a well-designed CNN

is exploited to infer the land-cover label of the center pixel in the HSI cube. In this paper, spectral-spatial-based methods that employ CNNs as the basic structures are referred to as CNN-based methods.

Although CNN-based methods have shown superior performance in recent literatures [17], [18], CNN-based methods still show poor generalization capabilities on complex land-cover layouts [19]. Recent CNN-based methods strongly rely on the assumption that adjacent pixels have the same land-cover category [14], [19]. Based on this assumption, in CNN-based methods, hyperspectral images are fixedly cropped into plenty of HSI cubes for HSI classification [18], [20], which may cause two problems as shown in Fig. 1(b). Firstly, the land-cover label of a cropped HSI cube is coarsely determined by its center pixel for training. However, when cropped HSI cubes are located at or near the edges of land-cover regions, cropped HSI cubes can contain different land-cover categories of hyperspectral pixels [11]. The land-cover label of a cropped HSI cube cannot be simply determined by its center pixel. Secondly, the observed scenes may be gathered with different viewing perspectives in an actual imaging environment [21], [22]. The spatial land-cover distributions in the same scene may be changed owing to the different viewing perspectives or urban renewal [6]. An excellent HSI classification method should be robust to the changes of spatial land-cover distributions. However, in CNN-based methods, after the cropping operation, the spatial land-cover distribution of a cropped HSI cube is fixed and has less diversity. Training on cropped HSI cubes can make the CNN-based methods overfit the spatial land-cover distributions of training samples [19].

In this paper, an end-to-end *fully convolutional segmentation network* (FCSN) is proposed for HSI classification. Different from coarse labels of HSI cubes in CNN-based methods, a fine label style is proposed to label each hyperspectral pixel of HSI cubes. When all hyperspectral pixels of HSI cubes are labeled finely, spatial land-cover distributions in HSI cubes can be easily obtained. First, different from the cropping operation in CNN-based methods, a HSI cube generation method is proposed to generate sufficient finely labeled HSI cubes for training. The proposed HSI cube generation method can randomly select and arrange training pixels in a horizontal or zig-zag mode to generate finely labeled HSI cubes, which improves the diversity of spatial land-cover distributions. Then, an end-to-end FCSN is proposed to explore both spectral signatures and spatial land-cover distributions of finely labeled HSI cubes for training. The proposed FCSN is a pixel-wise method, where the land-cover labels of all hyperspectral pixels in a HSI cube can be identified simultaneously. The proposed FCSN is composed of several simple residual blocks. Residual connection [23] is utilized to facilitate the convergence of FCSN. With the proposed HSI cube generation method, FCSN shows the superior generalization capability to the changes of spatial land-cover distributions.

The main contributions are as follows:

- To fully explore spatial land-cover distributions of HSI cubes, a fine label style is proposed to label all hyperspectral pixels of HSI cubes. With fine labels, the detailed

spatial land-cover distributions of the HSI can be easily obtained and exploited for HSI classification.

- To improve the diversity of spatial land-covers distributions to HSI cubes, a HSI cube generation method is proposed to generate various HSI cubes with fine labels. The proposed HSI cube generation method focuses on fully exploring training pixels to generate HSI cubes.
- To alleviate the problem of poor generalization capability, an end-to-end FCSN is proposed to classify hyperspectral data. FCSN does not rely on the assumption that adjacent pixels share the same land-cover category. FCSN classifies all hyperspectral pixels of a HSI cube simultaneously and shows high computational efficiency during testing.

Rest sections are organized as follows. First, plentiful works related to HSI classification are reviewed in Section II. Second, the motivation of FCSN is introduced in Section III. Third, Section IV introduces the details of FCSN. Then, experimental results are reported in Section V. Finally, Section VII summarizes this paper.

II. RELATED WORK

HSI classification has been extensively studied [14], [17]. Previous works usually utilize the superpixel segmentation method (*e.g.* simple linear iterative clustering [24], [25] or ultrametric contour map [26]) and the maximum a posterior framework [27] to explore the spatial information of HSIs. Recent works [7], [28] usually focus on designing various end-to-end deep networks suitable for HSI classification, which show superior performance [9], [17]. This paper mainly focuses on the generalization capabilities of deep learning-based methods. Recent deep learning-based methods are reviewed in this section.

A. Deep Learning-Based Methods

According to whether adjacent pixels are utilized, recent deep learning-based methods are simply divided into spectral-based methods and spectral-spatial-based methods.

Spectral-based methods mainly focus on designing various 1D deep networks to extract spectral features from 1D spectral signatures for HSI classification [10], [29]. Hu *et al.* [10] proposed a 1D convolutional network to directly explore the spectral signatures of pixels for HSI classification. Mou *et al.* [29] employed a recurrent neural network to extract spectral features based on the relationship between different spectral bands. Wu and Prasad [30] proposed to exploit both 1D convolutional network and recurrent neural network to extract contextual features from spectral signatures. However, spectral-based methods cannot explore the spatial information of hyperspectral data for HSI classification [13].

Spectral-spatial-based methods focus on exploiting both spectral signatures and spatial information of hyperspectral data for HSI classification [28], [31]. Previous spectral-spatial-based methods usually combine deep networks and hand-crafted features for HSI classification [28], [32]. Hand-crafted feature descriptors are utilized to extract features first and then deep networks are employed to further extract deep abstract features from extracted hand-crafted features. Chen *et al.* [28]

first employed gabor filters to extract the spatial texture and edge features from HSI. And then extracted spatial features were fed to a CNN for classification. Aptoula *et al.* [32] first employed the attribute profiles to extract geometric properties. Then, a five layer CNN was proposed to extract deep features from geometric properties. However, hand-crafted feature descriptors cannot flexibly extract various attributes to fully represent hyperspectral data [33]–[35].

Recent spectral-spatial-based methods are devoted to designing end-to-end deep networks for HSI classification [13], [36]. Chen *et al.* [37] proposed a 3D CNN combined with the L2 regularization. Fang *et al.* [12] proposed a multibias module which incorporates multibias into a activation layer to obtain different land-cover response maps. The multibias module was utilized to replace the traditional convolution for the end-to-end training. Han *et al.* [38] proposed a part-based convolutional network to extract discriminative features from object parts. Cheng *et al.* [21] designed a rotation-invariant layer to alleviate the image rotation problem. Yao *et al.* [39] proposed a stacked discriminative sparse autoencoder to improve the discrimination of features. Han *et al.* [22] proposed a deep Boltzmann machine to represent the spatial structural information of images. Zhong *et al.* [13] utilized residual connections [23] and batch normalization to construct a spectral-spatial residual network. Yu *et al.* [16] embedded hash functions into a CNN to enhance the discrimination of features. Paoletti *et al.* [9] proposed a pyramidal residual unit to build a deep convolutional network. Liu *et al.* [20] proposed a siamese network to mitigate the problem of insufficient labeled pixels. However, most spectral-spatial-based methods strongly rely on the assumption that adjacent pixels have the same land-cover category [14], [19] to preprocess hyperspectral data and design deep networks for HSI classification [17], [40], [41]. Spatial land-cover distributions of hyperspectral data cannot be fully explored in these methods.

B. HSI Classification With Segmentation Networks

The HSI classification task focuses on identifying the label of each pixel, which is similar to the common segmentation task [42]. Recently, some works [42], [43] have attempted to migrate existing segmentation methods to the HSI classification task. Jiao *et al.* [43] leveraged a fully convolutional network pre-trained on the ImageNet to extract potential spatial structural features, which were further fused with spectral features for HSI classification. Niu *et al.* [42] employed the DeepLab V2 as a spatial feature extractor, where the first three principal components of HSI were utilized as the input and the first principal component of HSI was utilized as the output for training. However, in these methods, segmentation networks are only utilized as spatial feature extractors. In addition, Lee and Kwon [44] also built a fully convolutional network for HSI classification, where three convolutional layers with different size kernels are added to the top of CNN to capture contextual information. However, the fully convolutional network proposed in the literature [44] is still a typical CNN-based method, which takes a cropped HSI cube as input and the label of the center pixel in the cube as output.

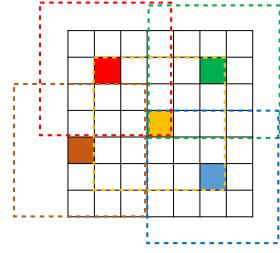


Fig. 2. Each block represents a hyperspectral pixel. Pixels of a HSI is randomly divided to training pixels (color blocks) and testing pixels (colorless blocks). In CNN-based methods, training HSI cubes (in dotted boxes) and testing HSI cubes have the similar spatial land-cover distributions.

III. MOTIVATION

Recent works usually focus on designing new deep networks to improve performance on public HSI databases [17]. Little attention is paid to two important issues about the generalization capabilities: one of which is whether designed deep convolutional networks are robust to the changes of spatial land-cover distributions, and the other is whether classification results are accurate at the edges of land-cover regions. In practical applications, spatial land-cover distributions of HSIs are variable and complicated due to changes of imaging perspective or urban renewal. An elaborate HSI classification method should be able to identify labels of hyperspectral pixels under different spatial land-cover distributions [19]. In addition, the purpose of HSI classification is to identify the land-cover category of each pixel in HSIs. In most literatures [19], [37], the performance of CNN-based methods is only evaluated in several isolated land-cover regions, not in the entire hyperspectral image. The ground truth on the existing HSI database is only covered several isolated land-cover regions [17]. Black land-cover regions in Fig. 1(b) are uncovered by the ground truth. On existing HSI databases, the performance of CNN-based methods is difficult to be measured in land-cover regions uncovered by the ground truth. Classification results on existing HSI databases cannot reflect the performance of CNN-based methods at the edges of land-cover regions.

To explore generalization capabilities of CNN-based methods, training samples and testing samples should have different spatial land-cover distributions [19]. However, in most CNN-based methods, due to that training and testing HSI cubes are randomly cropped from the same HSI [12], [13], training HSI cubes and testing HSI cubes usually share similar spatial land-cover distributions [19]. As shown in Fig. 2, pixels of a hyperspectral image are first randomly divided to training pixels and testing pixels. Then, in CNN-based methods, HSI cubes centered on training pixels are copped for training and HSI cubes centered on testing pixels are cropped for testing. There is always an adjacent training HSI cube whose spatial land-cover distribution is similar to the testing HSI cube [19]. Furthermore, in existing HSI databases, different land-cover regions are isolated by unlabeled regions [17]. The performance of CNN-based methods at the edges of land-cover regions cannot be evaluated in existing HSI databases. Therefore, due to similar spatial land-cover distributions and

isolated land-cover regions [12], [13], [17], existing HSI databases cannot be utilized to directly explore the generalization capabilities of CNN-based methods.

In this paper, based on the existing HSI database, an augmented HSI database with different spatial land-cover distributions is reconstructed to explore generalization capabilities of CNN-based methods. Details of experimental settings and the augmented database are introduced in Section V-B. From experimental results in Section V-B, it can be known that CNN-based methods have poor generalization capabilities to the changes of spatial land-cover distributions. The visual classification maps show that CNN-based methods perform poorly at edges of land-cover regions. The reason for poor generalization capabilities is that CNN-based methods cannot effectively exploit spatial land-cover distributions of HSI cubes for HSI classification. On the one hand, each HSI cube is coarsely labeled with the category of its center pixel. However, the label of the center pixel cannot reflect the complex spatial land-cover distribution of the HSI cube. Especially, as the spatial size of HSI cubes increases, spatial land-cover distributions of HSI cubes become more complicated. On the other hand, spatial information of HSI cubes can be lost in the fully connected layer, the flattening layer or the global pooling layer of CNN-based methods.

IV. PROPOSED METHOD

In this paper, an end-to-end FCSN is proposed for HSI classification. First, all pixels in HSI cubes should be labeled finely to provide detailed spatial land-cover distributions of HSI cubes. Then, the proposed FCSN explores both spectral signatures and spatial land-cover distributions to classify all pixels of a HSI cube simultaneously. Crucially, finely labeling all pixels of HSI cubes does not mean manually making ground truth for all pixels of existing HSI databases. In this paper, a HSI cube generation method is proposed to fully exploit existing training pixels to generate HSI cubes with different spatial land-cover distributions. Since the label of each pixel in training pixels is known, labels of all pixels in generated HSI cubes can be obtained for free.

Let a HSI be $\mathcal{X} \in \mathbb{R}^{H \times W \times C}$, where the spatial size of \mathcal{X} is $H \times W$ and the spectral band number of \mathcal{X} is C . The HSI \mathcal{X} totally contains $H \times W$ pixels and the spectral signatures of i -th pixel in \mathcal{X} is defined as $x_i \in \mathbb{R}^C$. HSI \mathcal{X} is assumed to contain K land-cover categories and let $\mathcal{Y} = \{1, 2, \dots, K\}$ represent the set of land-cover labels. The purpose of HSI classification is to estimate the land-cover label $\tilde{y}_i \in \mathcal{Y}$ for pixel x_i . In this paper, a fully convolutional segmentation network is proposed to classify HSI. The proposed FCSN takes a HSI cube as input and directly outputs a land-cover map of the HSI cube. Details of FCSN are introduced as follows. First, to clearly understand the proposed FCSN, differences between CNN-based methods and FCSN are described in Section IV-A. Second, the proposed HSI cube generation method is introduced in Section IV-B. Third, Section IV-C introduces the proposed FCSN in detail. Finally, the details of training and testing are introduced in Section IV-D.

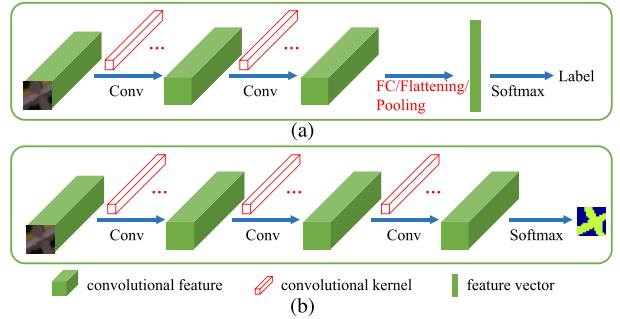


Fig. 3. (a) Brief structures of CNNs. (b) Brief structures of segmentation networks. ‘Conv’ represents the convolutional layer and ‘FC’ represents the fully connected layer.

A. CNN-Based Methods and the Proposed FCSN

This subsection introduces the differences between existing CNN-based methods and the proposed FCSN in terms of the styles of labeling HSI cubes and network structures.

1) Labeling HSI Cubes: Inspired by the superior performance of CNNs in image classification [23], [45], many works have attempted to migrate CNNs to HSI classification. Due to special structural designs of CNNs, the inputs of most CNN-based methods must be HSI cubes. Fortunately, adjacent pixels in hyperspectral data usually share the same land-cover category with high probability [14]. Therefore, to classify a hyperspectral pixel, its corresponding adjacent pixels are explored in CNN-based methods. During the training process, the label of the center pixel is used as the label of the HSI cube for training. During the test, the trained CNN is utilized to infer the label of the input HSI cube and serve as the label for the center pixel of HSI cube. However, as shown in Fig. 1, in cropped HSI cubes, adjacent pixels do not always share the same land-cover category. Therefore, the label of a hyperspectral pixel and the label of its corresponding HSI cube cannot be coarsely equivalent to each other. Different from existing CNN-based methods, in the proposed FCSN, a fine label method is proposed for the HSI cube where each pixel in a HSI cube should be labeled.

2) Network Structures: The basic operations for designing a network usually contains convoluting, activating, normalizing, pooling, fully connecting, classifying, and so on [38], [39]. Basic operations are hierarchically connected with different styles (*e.g.* residual connection [23], dense connection [45]) to form different networks. As shown in Fig. 3, the structures between the CNNs and the segmentation networks are similar in general, but the details are different. In CNN-based methods, a fully connected layer (or a global pooling layer) is usually utilized to convert convolutional features to a vector. The spatial information of HSI cubes can be lost in these layers [42], [43]. Different from CNN-based methods, FCSN focuses on classifying all pixels of HSI cubes where the spatial information of HSI cubes is the vital clue. Therefore, the proposed FCSN does not employ the fully connected layer, flattening layer and global pooling layer as basic operations.

CNN-based methods focus on classifying the center pixel of a HSI cube, while FCSN focuses on classifying all pixels

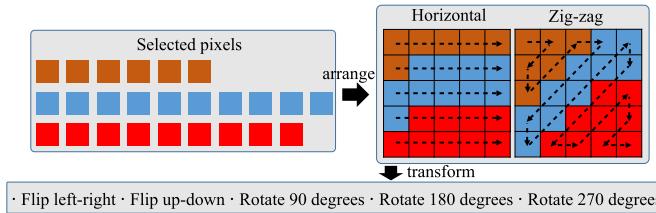


Fig. 4. A HSI cube generation method for generating finely labeled HSI cubes. Small squares of different colors represent pixels of different land-cover categories. First, several labeled pixels are randomly selected from the training pixels. Second, selected pixels are arranged in a horizontal mode or a zig-zag mode to form a HSI cube. Finally, flipping and rotating operations are employed to increase sample diversity.

of a HSI cube. From the perspective of outputs, a significant difference between CNN-based methods and the proposed FCSN is that CNN-based methods take a HSI cube as input and the label of center pixel as the output, while FCSN takes a HSI cube as input and labels of all pixels as outputs.

B. HSI Cube Generation

Similar to CNN-based methods [18], training FCSN also requires plenty of labeled hyperspectral data. In particular, each pixel of hyperspectral data should be finely labeled for training FCSN. However, in existing public HSI databases, the ground truth usually cannot cover all pixels [19]. Therefore, it is knotty to directly crop HSI cubes from existing HSI databases to train FCSN, because cropped HSI cubes often contain unlabeled pixels. It is very difficult to calibrate land-cover labels for all hyperspectral pixels in HSI databases. In this paper, a novel HSI cube generation method is proposed to utilize existing labeled pixels to generate finely labeled HSI cubes. The main idea of the proposed HSI cube generation method is to randomly select pixels from the labeled pixels and then arrange selected pixels into HSI cubes. A key principle for HSI cube generation is that spectral signatures and spatial land-cover distributions of generated HSI cubes should be as realistic as possible. Since pixels of generated HSI cubes are from existing HSI database, spectral signatures of generated HSI cubes are real. Therefore, only the spatial land-cover distributions of generated HSI cubes need to be as realistic as possible. In the proposed HSI cube generation method, spatial land-cover distributions of generated HSI cubes are determined by pixel arrangement types. In this paper, two pixel arrangement types are proposed to simulate realistic spatial land-cover distribution. The first type of pixel arrangement is to horizontally arrange pixels. The second type of pixel arrangement is to arrange pixels in a zig-zag mode. Horizontal and zig-zag arrangement types are shown in Fig. 4. Finally, to increase the diversity of spatial land-cover distribution, generated HSI cubes are flipped left-right, flipped up-down, rotated 90 degrees, rotated 180 degrees and rotated 270 degrees, respectively.

To mitigate the problem of insufficient labeled HSI pixels, mixture-based pixel generation [46] is employed to increase the number of labeled pixels. Pixels of the same land-cover category have similar spectral signatures. The mixture of two

Algorithm 1 Generating a HSI Cube and Its Corresponding Ground-Truth Map

Input: Training pixels x_i and corresponding labels y_i , the maximum number M of land-cover categories contained in a HSI cube, and the spatial size s of the HSI cube.

Output: A HSI cube and corresponding ground-truth map.

- 1: **Begin**
 - 2: Randomly select M_c ($1 < M_c < M$) land-cover categories, where generated HSI cubes are composed of these selected land-cover categories;
 - 3: Randomly select N_i ($1 < N_i < s * s$) HSI pixels from i -th selected land-cover category ($i = 1, 2, \dots, M_c$), where $\sum N_i = s * s$;
 - 4: Calculate the affinity with Eq. 1;
 - 5: Mix hyperspectral pixels with Eq. 2;
 - 6: Arrange pixels and labels to form a cube and corresponding ground-truth map with horizontal and zig-zag arrangement types.
 - 7: Transform generated HSI cubes with flipping and rotating operations.
 - 8: **End**
-

pixels from the same land-cover category should also belong this land-cover category [37], [46]. Two pixels (x_i and x_j) are mixed based on the affinity between x_i and x_j . The affinity can be formulated as follows:

$$\alpha_{ij} = e^{(-||x_i - x_j||^2 / 2\sigma^2)}, \quad (1)$$

where α_{ij} represents the affinity between x_i and x_j , and σ is set to 1 in this paper. The mixture of x_i and x_j are formulated as follows:

$$\bar{x}_i = (1 - \beta * \alpha_{ij}) * x_i + \beta * \alpha_{ij} * x_j, \quad (2)$$

where \bar{x}_i is the mixture of x_i and x_j , β is a random binary variable to decide whether to mix two pixels. To clearly understand the proposed HSI cube generation method, the detailed process is summarized in Algorithm 1.

C. Details of FCSN

In this paper, an end-to-end FCSN is proposed for HSI classification. Different from CNN-based methods, which only identify the land-cover label for the center pixel in a HSI cube [14], [17], the proposed FCSN simultaneously identifies land-cover labels for all pixels in a HSI cube. Recently, the residual connection has shown superior performance in HSI classification [9], [13], [33]. In this paper, the residual connection is also employed to construct the FCSN for HSI classification. First, the residual connection is exploited to build a residual block. Then, several residual blocks are stacked in series to form the proposed FCSN. The structures of FCSN are introduced as follows.

1) *Residual Blocks:* Details of residual blocks are shown in Fig. 5. Following the residual designing of the literature [23], a residual block is simply composed of two convolutional layers, two *batch normalization* (BN) layers and two *rectified linear unit* (ReLU) layers. Let $X^{in} \in \mathbb{R}^{s \times s \times c}$ represent the

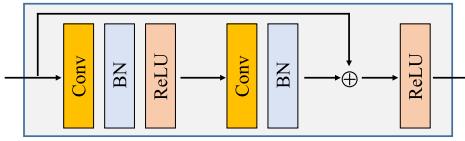


Fig. 5. The structure of a residual block. ‘BN’ represents the batch normalization. ‘Conv’ represents the convolution operation. ‘ReLU’ represents the rectified linear unit. \oplus represents the residual connection.

convolutional features, where the spatial size of X^{in} is $s \times s$ and the channel number of X^{in} is c . Extracting residual feature X^{out} from X^{in} is utilized as an example to introduce the residual block. First, a convolutional layer combined with a BN layer and a ReLU layer is employed to extract nonlinear feature \hat{X} from X^{in} . Second, extracted feature \hat{X} is fed to another convolutional layer combined with a BN layer for convolving and batch normalizing to generate $BN(Conv(\hat{X}))$ and X^{in} . Then, the residual connection is employed to connect $BN(Conv(\hat{X}))$ and X^{in} . Finally, a ReLU layer is utilized for activating. In all convolutional layers of a residual block, the padding operation is exploited to keep the spatial size of features unchanged. Spatial structure information of HSI cubes can be effectively maintained by the padding operation. The residual block can be formulated as follows:

$$X^{out} = ReLU(F(X^{in}) + X^{in}), \quad (3)$$

$$F(X^{in}) = BN(Conv(\hat{X}; W^{(2)}, b^{(2)})), \quad (4)$$

$$\hat{X} = ReLU(BN(Conv(X^{in}; W^{(1)}, b^{(1)}))), \quad (5)$$

where $ReLU(\cdot)$ represents a ReLU layer, $BN(\cdot)$ represents a BN layer, $Conv(\cdot)$ represents a convolutional layer, $F(\cdot)$ is composed of several convolution, BN and ReLU layers, $W^{(1)}, b^{(1)}, W^{(2)}, b^{(2)}$ are weight and bias parameters of convolutional layers. Kernel number of convolutional layers is set to 64 and spatial size of convolutional kernels is set to 3×3 in the residual block. The size of X^{out} and \hat{X} is $s \times s \times 64$.

2) *FCSN*: As shown in Fig. 6, the proposed FCSN is simply composed of multiple residual blocks. First, a convolutional layer combined with a BN layer and a ReLU layer is utilized to reduce the spectral dimension of HSI data. Then, five residual blocks are connected in series to extract spectral-spatial features. Finally, a softmax layer are employed to classify each pixel. In the proposed FCSN, all convolutional layers except the last convolutional layer are configured with 64 3×3 kernels. The last convolutional layer is configured with K 3×3 kernels, where K is the number of land-cover categories.

In FCSN, the size of input HSI cube X_i is $s \times s \times C$ and the size of features \hat{X}_i from last convolutional layer is $s \times s \times K$. Each spatial position of convolutional feature \hat{X}_i is one-to-one corresponding to each hyperspectral pixel of HSI cube X_i . To obtain the land-cover label of each pixel of X_i , a softmax layer is employed to normalize the value of \hat{X}_i to [0,1]. CNN-based methods usually flatten the feature \hat{X}_i into a vector, and then feed the flattened vector into a softmax. Different from CNN-based methods, the proposed FCSN sequentially inputs each position of \hat{X}_i into a softmax to obtain the confidence that pixels belong to different land-cover labels. The j -th spatial

position of \hat{X}_i is defined as $\hat{X}_i^j \in \mathbb{R}^{1 \times 1 \times K}$. The predicted confidence $\tilde{y}_i^{j,k}$ that j -th pixel X_i^j of X_i belongs to k -th land-cover category can be calculated as follows.

$$\tilde{y}_i^{j,k} = \frac{e^{\hat{X}_i^{j,k}}}{\sum_t^K e^{\hat{X}_i^{j,t}}}. \quad (6)$$

To train the proposed FCSN, the cross entropy cost is employed as the loss function. The loss of a batch in the training phase can be calculated as follows.

$$Loss = -\frac{1}{M} \sum_{m=1}^M \sum_{j=1}^{s \times s} \sum_{k=1}^K y_i^{j,k} \log \tilde{y}_i^{j,k}, \quad (7)$$

where M is the number of training batch size and $y_i^{j,k}$ is the real confidence that X_i^j belongs to k -th land-cover label.

D. Training and Testing

The proposed HSI cube generation method can generate millions of different finely labeled HSI cubes. To fully explore different HSI cubes for HSI classification, the proposed HSI cube generation method is integrated with the training process. In each iteration, the proposed HSI cube generation method is first utilized to generate HSI cubes, which are then fed to FCSN for training.

To evaluate the performance of FCSN, original HSI database is first cropped into HSI cubes without overlap. The spatial size of cropped HSI cubes for testing is simply set to the same size as the training set. Then, cropped HSI cubes are fed to FCSN to obtain land-cover labels of pixels. Due to that CNN-based methods should crop original HSI database in an overlapping manner for testing, the testing time of the proposed FCSN is significantly less than CNN-based methods.

V. EXPERIMENTAL SETTINGS

A. Databases and Evaluation Indexes

Three real HSI databases and a simulated HSI database are employed to explore the generalization capability of FCSN.

*Indian Pines (IP) Database*¹: This HSI is acquired by a AVIRIS sensor and contains 145×145 pixels, where each pixel consists of 224 spectral bands. The IP database contains 16 land-cover categories, including agriculture, forest and vegetation. On the IP database, number of labeled pixels in different categories varies greatly. For example, ‘Oats’ has only 20 labeled pixels, whereas ‘Soybean-mintill’ has 2455 labeled pixels.

*Pavia University (PU) Database*¹: The PU database is collected by a ROSIS sensor and contains 610×340 pixels, where pixels have 103 bands. The PU database contains 9 common land-cover categories in campus, e.g. ‘Meadows’, ‘Trees’ and ‘Self-blocking bricks’. There are plenty of labeled pixels on the PU database, and the number of labeled pixels in each land-cover category exceeds 900.

¹http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes

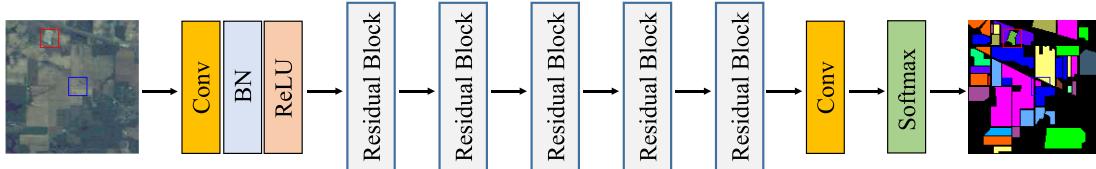


Fig. 6. Detailed structures of FCSN. The proposed FCSN is generated by simply stacking multiple residual blocks.

Mississippi Gulfport (MG) Database² [47], [48]: The MG database contains 325×220 pixels, where each pixel has 64 spectral bands. The observed scene on this database is also the campus. There are 11 common land-cover categories, such as ‘Trees’, ‘Grass ground surface’, ‘Water’, and ‘Buildings’. In each land-cover category, the number of labeled pixels exceeds 180.

Simulated Database³ [49]: This simulated database is generated by five real endmembers. Five endmembers with 162 spectral bands are mixed with the generalized bilinear mixing to obtain hyperspectral pixels. This simulated database contains total 40,000 hyperspectral pixels and the spatial size of the simulated database is 200×200 . The number of land-cover categories is 5.

This paper explores the generalization capabilities of the proposed FCSN and compared methods from two aspects: one is the generalization capabilities of classification methods under different spatial land-cover distributions, and the other is the generalization capabilities of classification methods at the edges of land-cover regions.

To explore generalization capabilities of different methods under different spatial land-cover distributions, training samples and testing samples should have different spatial land-cover distributions. A feasible approach is to train on one HSI database and then test on another HSI database with the same land-cover categories but with different spatial land-cover distributions. However, existing HSI databases usually contain only one image and land-cover categories between different databases are usually different. It is difficult to utilize existing databases to directly assess generalization capabilities. In this paper, labeled pixels of the existing HSI database are rearranged to reconstruct an augmented hyperspectral image, which has different spatial land-cover distributions from the existing databases, but the same land-cover categories. The augmented database can be utilized to explore the generalization capabilities under different spatial land-cover distributions. The IP, PU and MG databases are employed to reconstruct new augmented databases for experiments. Details of reconstructing augmented databases are introduced in Section V-B. Another feasible approach is to divide an existing hyperspectral image into two small hyperspectral images, one for training and the other for testing. However, existing HSI databases contain many land-cover categories and hyperspectral pixels of the same land-cover category are close in spatial position. Directly dividing an existing hyperspectral image will result in that two small hyperspectral images

contain different land-cover categories. The simulated database contains only 5 land-cover categories. After dividing, two small images of the simulated database contain the same land-cover categories and different spatial land-cover distributions. Therefore, the simulated database is also employed to conduct experiments.

To explore generalization capabilities of different methods at the edges of land-cover regions, labeled land-cover regions should be spatially continuous. However, on existing HSI databases, land-cover regions of different categories are usually separated by pixels of the unknown land-cover category. It is difficult to test the performance of classification methods on pixels of unknown land-cover category. To explore generalization capabilities at the edges of land-cover regions, all pixels on HSI databases should be labeled finely. On the augmented databases and the simulated database, each pixel is labeled finely. Therefore, the generalization capabilities at the edges of land-cover regions can be explored on augmented databases.

In this paper, *overall accuracy* (OA), *accuracy in each category* (AC), *average accuracy* (AA), and kappa [11], [19] are utilized to quantitatively evaluate the performance of FCSN. Since there is no quantitative evaluation index for evaluating the performance at land-cover region edges in HSI classification, this paper only conducts qualitative analysis based on visual classification maps.

B. Generalization Capabilities of Recent Methods

To comprehensively explore the generalization capabilities of recent methods, three different experimental conditions are set to evaluate recent methods, including testing on the original HSI database, testing on the rotated database, and testing on the augmented database with different spatial land-cover distribution. In this subsection, the IP database is employed to conduct experiments. First, a simple rotation operation is utilized to rotate testing samples at different degrees to explore the generalization capabilities of different methods. Then, an augmented IP database with fine labels is reconstructed based on the IP database. These methods are evaluated on both the original IP database and the augmented IP database to further explore the generalization capabilities.

Four recent deep learning-based methods are employed to conduct experiments, including *1D convolutional neural network* (1D CNN) [10], *2D convolutional neural network* (2D CNN) [37], *deformable HSI classification networks* (DHC-Net) [50], and *squeeze multi-bias network* (SMBN) [12]. 1D CNN is a spectral-based method and other methods are spectral-spatial-based methods. Details of experiments are as follows. Firstly, 10% of hyperspectral pixels in each category

²<https://github.com/GatorSense/MUUFLGulfport>

³<http://www.math.ucla.edu/~weizhu731/>

TABLE I
OAs (%) ON THE ORIGINAL IP DATABASE AND THE AUGMENTED IP DATABASE

Spatial Size	Testing on original IP				Testing on augmented IP			
	2-D CNN	DHCNet	SMBN	FCSN	2-D CNN	DHCNet	SMBN	FCSN
3 × 3	73.91	84.17	83.94	85.15	60.98	66.14	65.18	87.41
7 × 7	89.73	94.47	90.63	87.96	67.67	69.34	66.76	90.45
11 × 11	94.55	97.68	95.58	84.39	65.44	67.75	68.66	89.64
15 × 15	96.84	98.01	96.96	84.03	62.39	66.23	54.67	88.96
19 × 19	97.29	98.47	97.47	82.60	56.05	61.14	53.68	88.48

TABLE II
OAs (%) AT DIFFERENT ROTATION DEGREES

Degrees	1-D CNN	2-D CNN	DHCNet	SMBN	FCSN
0	83.59	91.29	95.53	91.46	87.96
90	83.59	83.51	80.89	72.88	87.63
180	83.59	83.35	81.36	74.38	87.81
270	83.59	85.37	80.54	80.74	87.85

of the original IP database are chosen as training pixels. Secondly, selected pixels are utilized to train 1D CNN and HSI cubes whose center pixels are training pixels are cropped from the original IP database to train other methods. Finally, these methods are respectively tested on different databases to explore the generalization capabilities.

1) *Generalization for Rotation*: In experiments, testing samples of the original IP database are rotated 0, 90, 180, and 270 degrees to verify the performance of different methods, respectively. Experimental results are shown in Table I. Due to that spectral-based methods do not explore spatial information from adjacent pixels for HSI classification, changing spatial land-cover distributions of testing samples does not affect the performance of 1D CNN. When the rotation degree is 0, as shown in Fig. 2, the testing samples and the training samples have similar spatial land-cover distributions. Spectral-spatial-based methods show superior performance when the rotation degree is 0. When testing samples are rotated, the spatial land-cover distributions of testing samples are different from that of training samples. After rotation, the performance of spectral-spatial-based methods is significantly reduced. It indicates that spectral-spatial-based methods are easy to overfit the spatial land-cover distributions of training samples. Spectral-spatial-based methods have poor generalization capabilities to the rotation of testing samples. OAs of FCSN at different rotation degrees are also shown in Table I. Compared with recent methods, OAs of FCSN at 90, 180, and 270 rotation degrees are significantly higher than spectral-spatial-based methods. The proposed FCSN has similar overall accuracies under different rotation degrees, which indicates that FCSN has the good generalization capability for rotation.

2) *Generalization on Original and Augmented Databases*: To further explore the generalization capabilities of recent methods, an augmented HSI database is manually reconstructed based on existing IP database. The augmented IP database is composed of the same land-cover categories as the IP database, but with different spatial land-cover distributions. The details of reconstructing the augmented database are

as follows. First, pixels of unlabeled regions on original IP database are removed. Second, as shown in Fig. 7, the black regions of ground truth map are manually labeled with several known land-cover categories. Finally, based on the manually labeled map, existing pixels corresponding to land-cover labels are randomly selected from original IP database to fill the position of removed pixels. The augmented IP database contains new spatial land-cover distributions, which can be utilized to explore the generalization capabilities of CNN-based methods to changes of spatial land-cover distributions. Furthermore, on the augmented IP database, different categories of land-cover regions are continuous in spatial distance. The performance of CNN-based methods at the edges of land-cover regions can also be explored.

To explore generalization capabilities, 2D CNN, DHCNet and SMBN are tested on the original IP database and the augmented IP database, respectively. Since the changes of spatial land-cover distributions do not affect the performance of spectral-based methods, 1D CNN is not employed to conduct experiments. Overall accuracies on the original IP database and the augmented IP database are reported in Table II. In spectral-spatial-based methods, the spatial size of HSI cubes is an important parameter, because the spatial size of HSI cubes determines the number of adjacent pixels. In experiments, the spatial size is set to 3 × 3, 7 × 7, 11 × 11, 15 × 15 and 19 × 19, respectively. As shown in Table II, overall accuracies of 2D CNN, DHCNet, and SMBN⁴ on the augmented IP database are much lower than overall accuracies on original IP database, indicating that these methods have poor robustness to spatial land-cover distribution changes. Especially, as the spatial size of HSI cubes becomes larger, overall accuracies on original IP database are getting higher, but overall accuracies on the augmented IP database are getting lower. Large spatial size of HSI cubes makes these methods overfit original IP database, which results in poor generalization capabilities. In addition, visual classification maps of 2D CNN, DHCNet and SMBN on the augmented IP database are displayed in Fig. 7. Using small size HSI cubes (3 × 3) as inputs can make classification results at the edges of land-cover regions better, but classification maps are not smooth. Using larger size HSI cubes (15 × 15, 19 × 19) makes classification maps smooth, but classification results are poor at edges of land-cover regions. An appropriate spatial size of HSI cubes is critical to the performance of CNN-based methods.

⁴Due to pooling layers in the network, SMBN cannot directly classify 3 × 3 and 7 × 7 cubes. To classify 3 × 3 and 7 × 7 cubes, pooling layers are removed from SMBN.

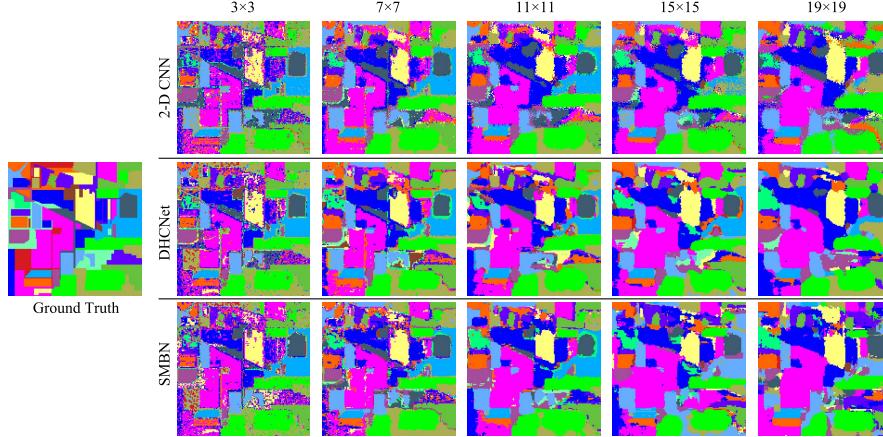


Fig. 7. Ground truth map of the augmented IP database and classification maps of 2D CNN, DHCNet, and SMBN tested on the augmented IP database.

TABLE III

OAs (%) OF FCSN CONFIGURED WITH DIFFERENT NUMBER OF KERNELS

Kernel Number	8	16	32	64	128	256
Original IP	79.45	82.02	85.21	86.40	84.89	84.63
Augmented IP	84.07	85.51	87.78	88.58	88.46	88.79

C. Detailed Settings of FCSN

To obtain optimal settings, FCSN with different settings are trained on the original IP database and tested on both the original IP database and the augmented IP database.

1) *Number of Convolutional Kernels:* Different numbers of kernels determine the number of weight parameters of FCSN. Too many weight parameters will make FCSN difficult to train, and too few weight parameters will make FCSN unable to fully represent HSIs [17], [33]. In this paper, different convolutional layers of FCSN are simply configured with the same setting, *e.g.* the same size and number of convolutional kernels. To explore the effect of kernel number on FCSN, number of residual blocks is set to 2, spatial size of HSI cubes is set to 7×7 , learning rate is set to 0.001, batch size is set to 100, and iteration epoch is set to 500. Then, the number of convolutional kernels is set to 8, 16, 32, 64, 128 and 256 for experiments, respectively. Overall accuracies of different numbers of convolutional kernels are reported in Table III. When the number of convolutional kernels is set to 64, FCSN achieves an overall accuracy of 86.40% on the original IP database and an overall accuracy of 88.58% on the augmented IP database. Therefore, kernel number of each convolutional layer is set to 64 for following experiments.

In this paper, because the spatial land-cover distribution complexity of the augmented database is much lower than that of the original database, FCSN performs better on the augmented database than on the original database.

2) *Number of Residual Blocks:* The residual block is the core component of the proposed FCSN. Excessive number of residual blocks will also introduce too many weight parameters, making FCSN difficult to train. To explore a proper number of residual blocks, FCSN is configured with multiple

TABLE IV

OAs (%) OF FCSN CONFIGURED WITH DIFFERENT NUMBER OF RESIDUAL BLOCKS

Block Number	1	2	3	4	5	6
Original IP	85.27	86.40	86.30	86.56	87.96	87.05
Augmented IP	85.43	88.58	89.39	90.28	90.45	90.32

residual blocks (ranging from 1 to 6 residual blocks) for experiments. As shown in Table IV, FCSN configured with 5 residual blocks achieves an overall accuracy of 87.96% on the original IP database and an overall accuracy of 90.45% on the augmented IP database. In this paper, the number of residual blocks is set to 5.

3) *Spatial Size of HSI Cubes:* In this subsection, experiments are conducted to explore the performance of FCSN with different spatial size HSI cubes. OAs are reported in Table II. When the spatial size of HSI cubes is 7×7 , FCSN achieves the best performance on both the original and augmented IP databases. As the spatial size of HSI cubes increases, the spatial land-cover distributions of HSI cubes become more complicated. The performance of FCSN is reduced under large size HSI cubes, due to that the complicated spatial land-cover distributions of large size HSI cubes are difficult to be simulated by the proposed HSI cube generation method. In this paper, the spatial size of HSI cubes is set to 7×7 for experiments.

4) *Performance of the HSI Cube Generation Method:* In the proposed HSI cube generation method, two pixel arrangement modes (horizontal mode and zig-zag mode) are proposed to generate HSI cubes with different spatial land-cover distributions. To show the effectiveness of the horizontal mode and the zig-zag mode, a random pixel arrangement mode is utilized as the comparison. OAs of FCSN with different pixel arrangement modes are reported in Table V. Since spatial land-cover distributions generated by the random mode are disorderly and have a large difference from realistic spatial land-cover distributions, the performance of FCSN with the random mode is worse than the performance of FCSN with the horizontal mode or the zig-zag mode. Especially, unrealistic spatial land-cover distributions generated by the random mode

TABLE V
OAs (%) OF FCSN WITH DIFFERENT PIXEL ARRANGEMENT MODES

Modes	Ran.	Hor.	Z-Z	All	Hor. + Z-Z
Original IP	81.28	86.68	85.63	85.13	87.96
Augmented IP	84.40	89.64	88.10	86.56	90.45

TABLE VI
NUMBER OF TRAINING PIXELS AND TESTING PIXELS ON THE IP DATA-
BASE

Category	Color	Name	Train	Test ¹	Test ²
1		Alfalfa	5	41	321
2		Corn-notill	143	1285	1785
3		Corn-mintill	83	747	1114
4		Corn	24	213	237
5		Grass-pasture	49	434	747
6		Grass-trees	73	657	1024
7		Grass-pasture-mowed	3	25	1348
8		Hay-windrowed	48	430	935
9		Oats	2	18	1047
10		Soybean-notill	98	874	1253
11		Soybean-mintill	246	2209	2942
12		Soybean-clean	60	533	1422
13		Wheat	21	184	1419
14		Woods	127	1138	2479
15		Buildings-grass-trees	39	347	1738
16		Stone-steel-towers	10	83	1214
Total			1031	9218	21025

1. ‘Train’ represents that number of training pixels on original IP database.

2. ‘Test¹’ represents that number of testing pixels on original IP database.

3. ‘Test²’ represents that number of pixels on augmented IP database.

can even degrade the performance of FCSN when the random mode, the horizontal mode and the zig-zag mode are employed simultaneously. In this paper, the horizontal and zig-zag modes are employed to generate HSI cubes for training.

VI. EXPERIMENTAL RESULTS

A. Experiments on Real Databases

In this subsection, the IP, PU and MG databases are utilized to evaluate the proposed FCSN. Recent 1D CNN [10], 2D CNN [37], DHCNet [50], SMBN [12], and *deep feature fusion network* (DFFN) [18] are employed as compared methods. In this paper, spectral-spatial-based methods that use CNNs as basic structures are referred to as CNN-based methods. 2D CNN, DHCNet, SMBN, and DFFN are CNN-based methods. For fair comparison, the same training settings are set for FCSN and CNN-based methods. Spatial size of HSI cubes is set to 7×7 for CNN-based methods. Batch size is 100. Iteration epoch is set to 500. In addition, two recent non-deep learning methods, *local covariance matrix representation* (LCMR) [51] and *invariant attribute profiles* (IAP) [52], are also employed as compared methods to further demonstrate the performance of FCSN. In LCMR and IAP, the SVM⁵ is utilized for classification, where the log-Euclidean-based kernel is employed in LCMR and the radial basis function kernel is employed in IAP. Five-fold cross validation is utilized to find optimal SVM hyperparameters for IAP.

1) *IP Database*: To fully explore generalization capabilities, FCSN and compared methods are trained on the original IP database and then tested on the original IP database and the augmented IP database. In this paper, 10% of labeled

pixels are randomly chosen from each land-cover category for training. On the IP database, the number of training pixels and testing pixels is shown in Table VI. Quantitative results are reported in Table VII. Due to that 1D CNN is a spectral-based method which does not introduce spatial information for HSI classification, 1D CNN is robust to the changes of spatial land-cover distributions. In the augmented IP database, the performance of 1D CNN even surpasses the performance of CNN-based methods. For CNN-based methods, on the original IP database, due to that training HSI cubes and testing HSI cubes are directly cropped from the original HSI, training HSI cubes and testing HSI cubes have similar spatial land-cover distribution (as shown in Fig. 2). CNN-based methods are easy to overfit training HSI cubes and show superior performance on the original IP database. In the ‘Hay-w’ and ‘Wheat’ categories of the original IP database, DHCNet and DFFN can even achieve an overall accuracy of 100.00%. However, on the augmented IP database, the performance of CNN-based methods drops a lot. Due to the poor generalization capabilities, CNN-based methods cannot effectively classify HSI cubes containing new spatial land-cover distributions. As shown in Table VII, the number of training pixels has a tremendous influence on the generalization capabilities of CNN-based methods. In land-cover categories with few training pixels, training HSI cubes contain few types of spatial land-cover distributions. On the augmented IP database, CNN-based methods show poor performance in land-cover categories with few training samples. In the ‘Oats’ category of the augmented IP database, DHCNet only achieves an overall accuracy of 6.30%. In the ‘Alfalfa’ and ‘Grass-p-m’ categories of the augmented IP database, overall accuracies of ‘SMBN’ are 14.33% and 2.89%, respectively.

On the original IP databases, LCMR and IAP show very superior performance, surpassing deep learning-based methods. However, LCMR and IAP also have poor generalization capabilities on the augmented IP database. LCMR utilized the spectral similarity to select local neighboring pixels for extracting the covariance matrix representation. IAP utilized the isotropic filters, the superpixel segmentation and the histogram of oriented gradients to extract the spatial-frequency joint features. The spatial land-cover distributions of HSIs are not fully explored by LCMR and IAP. On the original IP database, the performance of FCSN is weaker than CNN-based methods. Classification results of FCSN on the original IP database indicate that realistic spatial land-cover distributions of observed scenes are very complex and the proposed FCSN and HSI cube generation method cannot fully represent the realistic spatial land-cover distribution. A preferable HSI cube generation method that can perfectly simulate the realistic spatial land-cover distributions still need to be explored. The proposed FCSN shows a robust generalization capability to different spatial land-cover distributions and achieves an overall accuracy of 90.45% on the augmented IP database. The proposed HSI cube generation method can alleviate the problem of insufficient training HSI cubes. The proposed FCSN achieves an overall accuracy of 86.92%, 89.69% and 68.29% in the ‘Alfalfa’, ‘Grass-p-m’ and ‘Oats’ categories of the augmented IP database, respectively.

⁵<https://www.csie.ntu.edu.tw/~cjlin/libsvm/>

TABLE VII
PERFORMANCE ON THE ORIGINAL IP DATABASE AND AUGMENTED IP DATABASE

Category	Testing on original IP database								Testing on augmented IP database							
	LCMR	IAP	1-DCNN	2-DCNN	DHCNet	SMBN	DFFN	FCSN	LCMR	IAP	1-DCNN	2-DCNN	DHCNet	SMBN	DFFN	FCSN
1	100.00	97.56	82.93	73.17	75.61	65.85	100.00	87.80	28.35	97.20	82.87	33.33	34.89	14.33	63.24	86.92
2	97.04	95.33	75.02	84.98	94.09	89.34	96.42	84.44	96.41	83.98	77.20	68.12	84.87	81.46	82.30	87.34
3	98.26	98.39	84.34	93.17	97.32	92.50	96.79	86.48	69.57	73.43	86.00	71.63	69.21	68.04	68.04	86.27
4	98.59	96.24	61.03	76.06	84.98	62.91	94.84	93.90	91.98	84.39	64.56	62.87	62.45	46.84	79.75	92.83
5	97.70	98.62	90.78	95.39	97.70	95.85	97.47	94.24	71.35	87.95	92.10	83.53	62.38	76.97	86.88	96.12
6	99.70	99.85	93.30	98.33	97.87	99.24	98.78	97.26	97.46	94.92	93.95	92.77	87.40	84.86	89.06	98.05
7	100.00	100.00	80.00	52.00	60.00	36.00	100.00	88.00	41.15	76.48	82.20	13.58	7.20	2.89	53.64	89.69
8	100.00	100.00	94.19	99.30	100.00	96.74	100.00	99.77	85.45	81.28	95.08	71.23	56.36	60.32	73.58	99.68
9	94.44	100.00	44.44	66.67	94.44	50.00	100.00	61.11	1.43	46.90	49.57	1.43	6.30	3.53	3.92	68.29
10	92.79	91.76	79.75	89.36	90.39	86.73	93.59	80.21	82.36	79.97	81.96	66.72	57.54	62.25	64.17	85.16
11	96.42	98.19	84.38	94.48	99.28	96.65	97.96	83.75	84.60	88.04	85.72	85.25	92.25	89.26	89.09	87.02
12	95.31	96.62	79.74	75.42	89.87	76.74	91.56	88.18	82.49	63.01	81.01	29.68	58.16	40.44	54.85	89.31
13	99.46	99.46	98.91	95.65	100.00	100.00	100.00	98.91	77.73	91.83	98.80	81.32	73.71	76.25	82.52	99.08
14	99.91	99.82	94.29	98.77	98.15	97.01	96.40	95.52	98.71	95.72	94.96	97.01	90.00	91.45	92.94	97.98
15	100.00	95.39	53.60	80.40	81.56	75.79	96.25	74.64	86.54	89.87	58.40	75.60	78.60	75.09	86.02	85.62
16	100.00	98.80	96.39	89.16	100.00	91.57	93.98	98.80	52.97	91.19	95.72	88.06	95.80	93.00	94.40	98.85
OA	97.44	97.42	83.59	91.29	95.53	91.46	96.69	87.96	74.18	83.58	83.50	68.57	69.75	67.62	75.88	90.45
AA	98.10	97.88	80.82	85.14	91.33	82.06	97.13	88.31	69.47	82.88	82.51	63.88	63.57	60.44	72.77	90.51
Kappa	97.08	97.05	81.29	90.05	94.89	90.24	96.23	86.30	71.95	82.12	82.08	65.78	66.96	64.67	73.77	89.63

Visual classification maps of the proposed FCSN and compared methods on the augmented IP database are shown in Fig. 8. For the ground truth of the augmented IP database, the visual classification map is smooth and edges of each land-cover region are regular shapes. Although the classification map of LCMR contains regular edges of land-cover regions, there is a lot of noise in many land-cover regions. The classification map of IAP is smoother than other methods. However, IAP shows poor performance at the edges of different land-cover regions. For the visual classification map of 1D CNN, the edges of land-cover regions are clear. However, because 1D CNN cannot exploit the spatial information of HSI, there is a lot of noise in land-cover regions. For the visual classification maps of 2D CNN, DHCNet, SMBN and DFFN, the edges of each land-cover region are not neat, since these methods strongly rely on the assumption that adjacent pixels have the same land-cover category. However, at or near the edges of land-cover regions, adjacent pixels in a cropped HSI cube usually have different categories. CNN-based methods cannot accurately identify the categories of these pixels. The classification map of FCSN is shown in Fig. 8(i). The visual classification map of FCSN has more neat edges of land-cover regions than that of CNN-based methods, and contains less noise than that of 1D CNN. Since the proposed FCSN can fully explore the spatial land-cover distributions of HSI, FCSN outperforms recent methods at edges of land-cover regions and has superior generalization capability. In particular, FCSN can effectively identify the labels of hyperspectral pixels in narrow land-cover regions.

2) PU Database: Similar to the IP database, existing labeled pixels of the PU database are also utilized to reconstruct an augmented PU database, where each pixel is finely labeled. 5% of pixels in each land-cover category on the original PU database are randomly chosen as training pixels. On the PU database, the number of training pixels and testing pixels is shown in Table VIII. The performance of FCSN and compared methods on the original PU database and the augmented PU database are reported in Table IX. With sufficient training pixels, compared methods achieve superior performance on the original PU database. Especially,

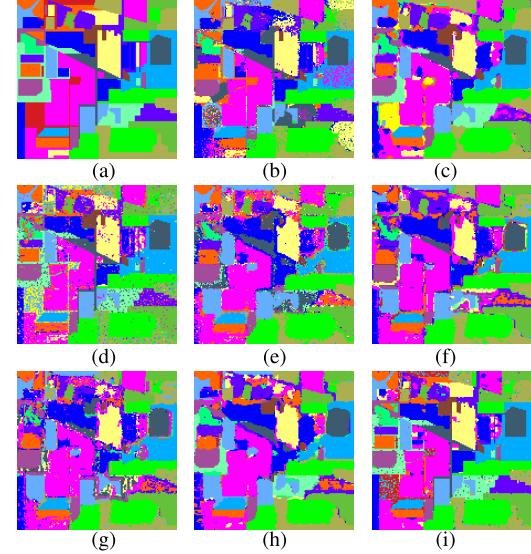


Fig. 8. Classification maps on the augmented IP database. (a) Ground truth. (b) LCMR. (c) IAP. (d) 1D CNN. (e) 2D CNN. (f) DHCNet. (g) SMBN. (h) DFFN. (i) The proposed FCSN. Best zoom-in view.

TABLE VIII
NUMBER OF TRAINING PIXELS AND TESTING PIXELS ON THE PU DATABASE

Category	Color	Name	Train	Test ¹	Test ²
1	Grey	Asphalt	332	6299	33493
2	Green	Meadows	933	17716	40270
3	Cyan	Gravel	105	1994	20607
4	Dark Green	Trees	154	2910	23305
5	Magenta	Painted metal sheets	68	1277	15340
6	Brown	Bare Soil	252	4777	23585
7	Purple	Bitumen	67	1263	18747
8	Red	Self-blocking bricks	185	3497	13352
9	Yellow	Shadows	48	899	18701
Total			2144	40632	207400

1. ‘Train’ represents that number of training pixels on original PU database.

2. ‘Test¹’ represents that number of testing pixels on original PU database.

3. ‘Test²’ represents that number of pixels on augmented PU database.

LCMR performs better than deep learning-based methods on the original PU database. However, on the augmented PU database, compared methods still show poor generalization

TABLE IX
PERFORMANCE ON THE ORIGINAL PU DATABASE AND AUGMENTED PU DATABASE

Category	Testing on original PU database								Testing on augmented PU database							
	LCMR	IAP	1-DCNN	2-DCNN	DHCNet	SMBN	DFFN	FCSN	LCMR	IAP	1-DCNN	2-DCNN	DHCNet	SMBN	DFFN	FCSN
1	98.92	99.30	91.03	94.59	98.51	98.78	99.00	95.30	89.51	87.59	91.11	84.61	90.81	86.81	88.67	98.69
2	99.75	99.86	96.79	99.84	99.97	99.75	99.92	92.96	80.67	88.87	96.86	81.16	54.82	49.29	45.09	96.53
3	98.29	91.62	75.03	59.98	93.23	93.03	95.64	78.44	89.15	80.48	75.60	38.93	57.20	73.13	75.81	95.98
4	99.18	98.08	92.65	96.46	98.93	97.29	98.93	96.32	95.99	95.06	92.71	76.82	91.68	80.38	88.93	98.79
5	99.53	99.92	99.61	99.84	100.00	100.00	100.00	99.53	96.65	99.07	99.56	98.81	95.32	94.86	97.14	99.82
6	99.92	99.71	68.14	93.34	99.62	98.95	99.87	95.50	85.32	86.50	68.79	86.16	84.42	82.33	89.01	98.66
7	99.37	97.94	89.07	88.84	97.94	93.59	97.70	93.75	98.00	75.00	89.30	77.74	84.96	81.24	86.84	99.02
8	99.14	95.94	86.62	99.00	98.46	98.03	98.97	95.94	76.56	76.91	87.24	74.00	75.37	69.56	71.09	97.93
9	96.22	97.66	99.33	98.67	100.00	100.00	100.00	99.19	80.02	99.34	86.28	93.93	94.28	94.67	99.95	
OA	99.38	98.78	90.19	95.62	99.11	98.67	99.35	93.79	89.34	86.19	89.28	78.60	78.95	76.59	78.87	98.19
AA	98.92	97.78	88.70	92.28	98.52	97.71	98.89	94.19	90.11	85.50	88.95	78.28	80.95	79.10	81.92	98.37
Kappa	99.18	98.38	86.86	94.18	98.82	98.24	99.13	91.86	87.88	84.21	87.71	75.62	76.09	73.46	76.07	97.93

TABLE X
NUMBER OF TRAINING PIXELS AND TESTING PIXELS ON THE MG DATA-
BASE

Category	Color	Name	Train	Test ¹	Test ²
1	[Color]	Trees	2325	20921	27575
2	[Color]	Grass ground surface	427	3843	4270
3	[Color]	Mixed ground surface	689	6193	7352
4	[Color]	Dirt and sand	183	1643	1826
5	[Color]	Road	669	6018	8742
6	[Color]	Water	47	419	2465
7	[Color]	Buildings	224	2009	3106
8	[Color]	Shadow of buildings	624	5616	10492
9	[Color]	Sidewalk	139	1246	1385
10	[Color]	Yellow curb	19	164	1656
11	[Color]	Cloth panels	27	242	2631
Total			5373	48314	71500

1. ‘Train’ represents that number of training pixels on original MG database.
2. ‘Test¹’ represents that number of testing pixels on original MG database.
3. ‘Test²’ represents that number of pixels on augmented MG database.

capabilities. On the augmented PU database, the proposed FCSN outperforms compared methods and achieves an overall accuracy of 98.19%.

Visual classification maps of the proposed FCSN and compared methods on the augmented PU database are shown in Fig. 9. Although the visual classification maps of LCMR and 1D CNN have relatively clear edges of land-cover regions, they also contain a lot of noise. In the classification maps of IAP and CNN-based methods, misclassified pixels are mainly concentrated at the edges of land-cover regions and in narrow land-cover regions. IAP and CNN-based methods still show the poor performance at the edges of land-cover regions on the augmented PU database. Compared with IAP and CNN-based methods, the classification map of FCSN contains neat edges in different land-cover regions. In the classification map of FCSN, misclassified pixels are scattered in different spatial locations. The proposed FCSN shows the superior performance at the edges of land-cover regions and in narrow land-cover regions.

3) *MG Database*: Although 75.09% of the pixels on the original MG database are labeled and different land-cover regions are relatively close, there are still 24.91% of the pixels that are not labeled. Different land-cover regions are isolated by the unlabeled pixels. To further explore generalization capabilities of FCSN and compared methods, an augmented MG database is reconstructed based on the original MG database. In each category, 10% of labeled pixels are randomly selected from the 1st row to the 160th row of the original

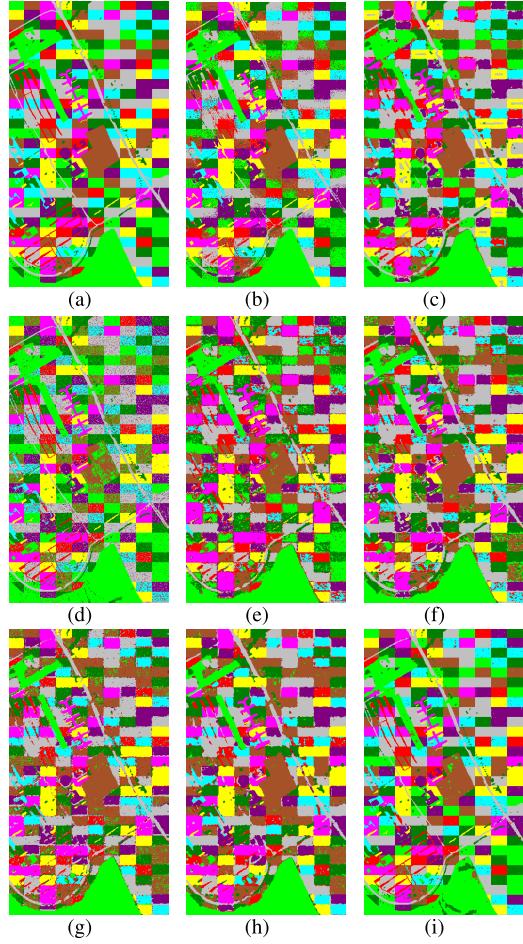


Fig. 9. Classification maps on the augmented PU database. (a) Ground truth. (b) LCMR. (c) IAP. (d) 1D CNN. (e) 2D CNN. (f) DHCNet. (g) SMBN. (h) DFFN. (i) The proposed FCSN. Best zoom-in view.

MG database as training pixels. The number of training pixels and testing pixels is shown in Table X. FCSN and compared methods are trained on the original MG database and tested on both original and augmented MG databases. Quantitative classification results of FCSN and compared methods are shown in Table XI. On the original MG database, although the OA, AA and Kappa of 1D CNN are lower than other methods, the accuracy of 1D CNN in the ‘Yellow curb’ category is much higher than that of LCMR and DHCNet. The land-cover region of the ‘Yellow curb’ category is very narrow. The accuracy

TABLE XI
PERFORMANCE ON THE ORIGINAL MG DATABASE AND AUGMENTED MG DATABASE

Category	Testing on original MG database								Testing on augmented MG database							
	LCMR	IAP	1-DCNN	2-DCNN	DHCNet	SMBN	DFFN	FCSN	LCMR	IAP	1-DCNN	2-DCNN	DHCNet	SMBN	DFFN	FCSN
1	98.48	95.46	94.70	97.69	97.45	98.12	97.64	83.05	97.59	94.02	94.75	94.76	95.41	94.56	94.21	89.90
2	78.48	82.85	78.64	83.40	85.62	82.76	85.50	83.86	71.55	76.07	78.64	65.81	67.96	74.24	73.11	86.84
3	81.37	64.83	67.19	74.75	76.88	77.36	79.26	78.02	70.57	58.05	67.30	67.66	70.87	67.76	72.99	73.37
4	74.98	68.84	71.85	85.38	84.50	81.98	77.98	94.45	48.74	67.80	71.85	65.99	69.99	73.82	73.38	74.86
5	86.01	88.50	78.11	93.75	92.00	95.07	92.43	91.75	81.18	86.34	78.11	77.25	73.78	82.58	80.28	92.68
6	91.17	86.87	82.40	87.34	96.14	94.21	98.71	97.85	64.26	42.19	82.03	34.12	41.95	43.25	54.40	97.65
7	81.63	72.52	66.23	82.13	85.00	84.95	89.25	93.19	71.15	54.57	65.61	66.39	64.07	67.77	71.12	94.40
8	79.75	89.03	88.46	93.35	93.22	93.91	94.78	95.13	79.73	66.02	88.76	79.11	84.14	78.78	73.34	93.71
9	54.74	52.49	46.57	61.23	69.89	64.04	65.63	68.85	46.35	31.55	46.57	42.45	49.03	46.50	52.49	57.47
10	53.05	15.85	72.13	30.05	42.62	21.31	31.15	88.89	66.55	20.11	71.01	6.28	5.25	1.39	2.17	83.70
11	94.21	37.60	49.07	43.49	67.66	61.71	71.75	97.00	88.18	4.10	47.09	7.11	50.36	51.58	59.83	96.01
OA	88.10	85.32	83.49	90.02	90.68	90.91	91.15	86.46	83.03	73.80	82.53	75.46	78.45	78.68	78.89	88.45
AA	79.44	68.62	72.31	75.69	81.00	77.77	80.37	88.37	71.44	54.62	71.98	55.17	61.16	62.02	64.30	85.51
Kappa	83.95	80.49	78.09	86.75	87.65	87.93	88.28	83.99	78.11	66.07	77.77	68.56	72.31	72.84	73.17	85.67

in the ‘Yellow curb’ category further shows that IAP and CNN-based methods are difficult to effectively identify narrow land-cover regions. On the original MG database, although the OA of FCSN is about 5% lower than that of CNN-based methods, the AA of FCSN is about 7% higher than that of compared methods. This is because compared methods overfit the ‘Trees’ category, and this land-cover category contains 43.30% of pixels in the testing pixels. In most land-cover categories of the original MG database, FCSN has higher accuracy than compared methods. Especially in the ‘Yellow curb’ category, the accuracy of FCSN is 46.27% higher than that of DHCNet. And the accuracy of FCSN is 25.25% higher than that of DFFN in the ‘Cloth panels’ category. On the augmented MG database, FCSN achieves the superior performance, which further verifies that the proposed FCSN has better generalization capability than compared methods to the changes of spatial land-cover distributions.

Visual classification maps of FCSN and compared methods on the augmented MG database are given in Fig. 10. As shown in Fig. 10(a), most pixels in this database belong to the ‘Trees’ category. Although compared methods achieve good classification results in the ‘Tree’ category, these methods perform worse than the proposed FCSN in most land-cover categories. On the augmented MG database, classification maps of compared methods contain many noises. As shown in Fig. 10(i), the classification map of FCSN is smoother than that of compared methods. FCSN can effectively identify the categories of pixels located in small and narrow land-cover regions. On the augmented MG database, the proposed FCSN still shows better generalization capability than CNN-based methods at the edges of each land-cover region.

B. Experiments on a Simulated Database

In this subsection, a simulated database (D) is also utilized to explore the generalization capability of FCSN and compared methods. Spatial size of the simulated database is 200×200 . As shown in Fig. 11(a), the simulated database is divided into two small hyperspectral images (D_1 and D_2). D_1 consists of pixels from the 1st row to the 20th row of the simulation database and D_2 consists of pixels from the 21st row to the 200th row of the simulation database. In experiments, 2% of labeled pixels in each category are randomly selected as training pixels. On the simulated database, the number

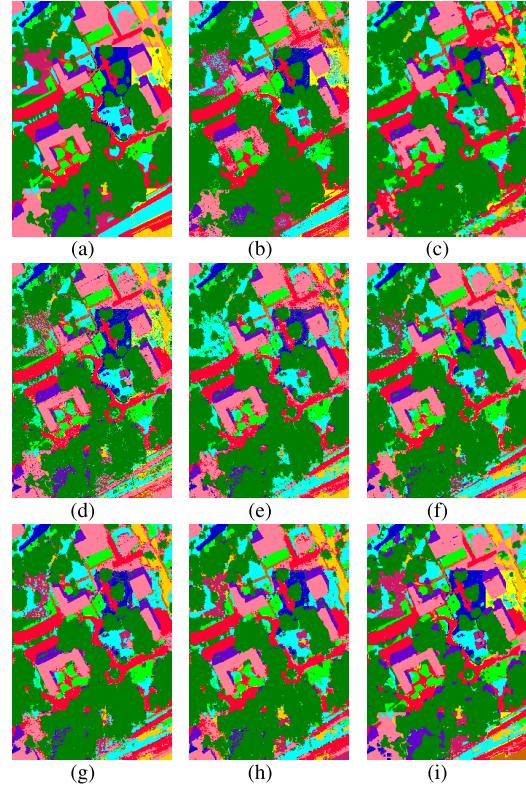


Fig. 10. Classification maps on the augmented MG database. (a) Ground truth. (b) LCMR. (c) IAP. (d) 1D CNN. (e) 2D CNN. (f) DHCNet. (g) SMBN. (h) DFFN. (i) The proposed FCSN. Best zoom-in view.

of training pixels and testing pixels is shown in Table XII. As shown in Table XIII, two experiments ($D_1 \rightarrow D_1$ and $D_1 \rightarrow D_2$) are conducted on the simulation database. In the first experiment, $D_1 \rightarrow D_1$ represents that hyperspectral pixels of each category on D_1 are randomly selected for training and the remaining pixels of D_1 for testing. 1D CNN achieves 99.98% overall accuracy in the $D_1 \rightarrow D_1$, which indicates that the spectral signatures of pixels between different land-cover categories are easy to distinguish. However, the performance of LCMR, IAP and CNN-based methods is lower than the performance of 1D CNN in the $D_1 \rightarrow D_1$. The introduction of spatial information reduces the performance of those methods. This shows that LCMR, IAP and CNN-based

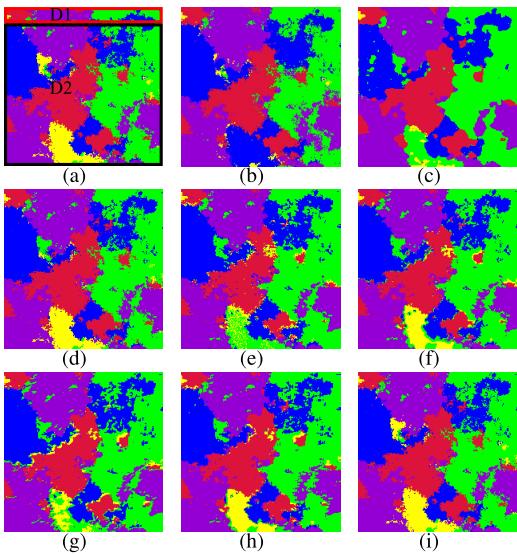


Fig. 11. Classification maps on the simulated database. (a) Ground truth. (b) LCMR. (c) IAP. (d) 1D CNN. (e) 2D CNN. (f) DHCNet. (g) SMBN. (h) DFFN. (i) The proposed FCSN. Best zoom-in view.

TABLE XII
NUMBER OF TRAINING PIXELS AND TESTING PIXELS ON THE SIMULATED DATABASE

Category	Color	Name	Train	Test ¹	Test ²
1	Red	Category 1	151	468	6922
2	Blue	Category 2	167	301	7870
3	Yellow	Category 3	47	7	2264
4	Green	Category 4	211	1172	9140
5	Purple	Category 5	226	1250	9804
Total			802	3198	36000

1. ‘Train’ represents that number of training pixels on $D1$.

2. ‘Test¹’ represents that number of testing pixels on $D1$.

3. ‘Test²’ represents that number of pixels on $D2$.

TABLE XIII
OAs (%) ON THE SIMULATED DATABASE

Methods	$D1 \rightarrow D1$	$D1 \rightarrow D2$
LCMR	99.84	89.48
IAP	93.06	86.51
1-D CNN	99.98	99.21
2-D CNN	96.06	87.93
SMBN	95.06	87.50
DFFN	95.22	90.56
DHCNet	97.37	91.64
FCSN	100.00	100.00

methods cannot effectively use spatial information for HSI classification. FCSN achieves the 100% overall accuracy better than compared methods in the $D1 \rightarrow D1$. In the second experiment, $D1 \rightarrow D2$ represents that FCSN and compared methods are trained on $D1$ and tested on $D2$. Training pixels in the $D1 \rightarrow D2$ are the same as the $D1 \rightarrow D1$. Compared with the $D1 \rightarrow D1$, the overall accuracies of compared methods in the $D1 \rightarrow D2$ are reduced. The main difference between $D1$ and $D2$ is the spatial land-cover distribution. The decrease in overall accuracies indicates that compared methods are not robust to different spatial land-cover distributions. The proposed FCSN still shows a robust generalization capability on the simulated database.

Visual classification maps of FCSN and compared methods tested on the entire simulated database are shown in Fig. 11. As shown in Fig. 11(a), the edges of different land-cover regions are irregular shapes, which complicate the spatial land-cover distributions near the edges of land-cover regions. On the simulated database, without atmospheric noises and instrument effects, the HSI classification task is less challenging. 1D CNN achieves the superior performance on the simulated database, which indicates that good classification results can be obtained by only using spectral signatures for HSI classification. However, with the introduction of spatial information, the performance of the compared methods becomes worse, which indicates that compared methods cannot effectively explore spatial land-cover distributions of HSIs. As shown in Fig. 11, the pixels misclassified by compared methods are mainly near the edges of land-cover regions. Especially, IAP shows the poor performance in small land-cover regions. The classification map of FCSN is shown in Fig. 11(i). The visual classification map of FCSN is the same as that of the ground truth. The proposed FCSN shows the superior performance at the edges of land-cover regions.

C. Training and Testing Time

To show computational efficiency of FCSN, the training and testing time is reported in Table XIV. All experiments of FCSN and compared methods are conducted on a workstation with the Ubuntu 14.04 system, Intel Core i7-5930K, NVIDIA Titan X, 64G RAM. Deep learning-based methods are implemented with the Tensorflow on the GPU. LCMR and IAP are implemented with the MATLAB on the CPU. In the testing phase, batch size is set to 100 for deep learning-based methods. During each iteration of the training phase, the proposed HSI cube generation method is utilized to generate HSI cubes with new spatial land-cover distributions for FCSN. As shown in Table XIV, training the proposed FCSN takes a lot of time, most of which is used by the proposed HSI cube generation method. During the testing phase, the proposed FCSN shows high computational efficiency, because that FCSN can classify all pixels in a HSI cube simultaneously.

D. Discussion

As shown in Tables VII, IX and XI, although the proposed FCSN performs better than CNN-based methods on the augmented databases, the performance of FCSN on the original databases is worse than CNN-based methods.

In CNN-based methods, HSI cubes centered on training pixels are directly copied from the original HSI as the training samples and HSI cubes centered on testing pixels are cropped as the testing samples. As shown in Fig. 2, on original databases, training samples and testing samples share similar spatial land-cover distributions. However, the testing samples on the augmented databases and the training samples on the original databases have different spatial land-cover distributions. CNN-based methods are easy to overfit training samples, which results in that CNN-based methods have high accuracies on original databases but perform poorly on the augmented

TABLE XIV
TRAINING AND TESTING TIME (SECONDS) ON HSI DATABASES

	IP		PU		MG		Simulated	
	Train	Test	Train	Test	Train	Test	Train	Test
LCMR	12.67	0.73	117.65	145.68	42.78	111.17	12.31	0.77
IAP	36.14	1.48	113.28	14.61	277.39	16.80	18.56	1.57
ID-CNN	18.92	0.54	42.20	6.64	88.31	1.78	15.33	0.70
2D-CNN	52.81	0.90	76.02	8.81	171.22	2.35	35.26	1.33
DHCNet	126.88	3.10	244.24	30.19	608.14	9.26	93.88	5.36
SMBN	217.24	2.65	419.76	25.49	1083.17	8.45	165.35	4.63
DFFN	146.84	1.84	310.16	20.57	828.79	6.95	126.25	3.56
FCSN	741.98	0.13	1449.47	0.48	3576.54	0.24	592.55	0.16

databases. CNN-based methods have poor generalization capabilities.

In the proposed FCSN, only training pixels are utilized to generate various training samples by the proposed HSI cube generation method. The training samples of FCSN are not cropped from the original HSI, which can avoid overfitting. However, realistic spatial land-cover distributions of observed scenes are very complicated. There is a certain degree of difference between the simulated training samples and the realistic testing samples in the spatial land-cover distributions, which results in that FCSN performs worse than CNN-based methods on original databases. The proposed FCSN focuses on fully exploiting the spatial land-cover distributions of HSIs for classification, and has good generalization capability. FCSN shows the superior performance on the augmented databases.

VII. CONCLUSION

This paper proposes an end-to-end FCSN for HSI classification. The proposed FCSN is simply composed of several residual blocks to facilitate converge. Experimental results demonstrate that FCSN with the proposed HSI cube generation method has robust generalization capability to changes of spatial land-cover distributions. Due to that FCSN can identify the label of each pixel in a HSI cube simultaneously, computational efficiency of the proposed FCSN is very high in the testing phase. However, experimental results on original databases show that realistic spatial land-cover distributions are very complicated and difficult to be simulated by the proposed HSI cube generation method. Exploring a superior HSI cube generation method to fully simulate realistic spatial land-cover distributions still requires further study. Further, FCSN is a supervised method that fails to utilize unlabeled pixels for HSI classification. In future, we will study semi-supervised segmentation networks to explore both labeled and unlabeled pixels for HSI classification.

REFERENCES

- [1] M. Zhang, W. Li, and Q. Du, “Diverse region-based CNN for hyperspectral image classification,” *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2623–2634, Jun. 2018.
- [2] W. Liu, X. Shen, B. Du, J. W. Tsang, W. Zhang, and X. Lin, “Hyperspectral imagery classification via stochastic HHSVMs,” *IEEE Trans. Image Process.*, vol. 28, no. 2, pp. 577–588, Feb. 2019.
- [3] F. Xiong, J. Zhou, and Y. Qian, “Material based object tracking in hyperspectral videos,” *IEEE Trans. Image Process.*, vol. 29, pp. 3719–3733, Jan. 2020.
- [4] X. Lu, Y. Yuan, and X. Zheng, “Joint dictionary learning for multispectral change detection,” *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 884–897, Apr. 2017.
- [5] R. A. Borsoi, T. Imbiriba, and J. C. M. Bermudez, “A data dependent multiscale model for hyperspectral unmixing with spectral variability,” *IEEE Trans. Image Process.*, vol. 29, pp. 3638–3651, Jan. 2020.
- [6] L. Drumetz, J. Chanussot, C. Jutten, W.-K. Ma, and A. Iwasaki, “Spectral variability aware blind hyperspectral image unmixing based on convex geometry,” *IEEE Trans. Image Process.*, vol. 29, pp. 4568–4582, Feb. 2020.
- [7] L. Mou and X. X. Zhu, “Learning to pay attention on spectral domain: A spectral attention module-based convolutional network for hyperspectral image classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 110–122, Jan. 2020.
- [8] J. Feng *et al.*, “CNN-based multilayer spatial–spectral feature fusion and sample augmentation with local and nonlocal constraints for hyperspectral image classification,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 4, pp. 1299–1313, Apr. 2019.
- [9] M. E. Paoletti, J. M. Haut, R. Fernandez-Beltran, J. Plaza, A. J. Plaza, and F. Pla, “Deep pyramidal residual networks for spectral–spatial hyperspectral image classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 740–754, Feb. 2019.
- [10] W. Hu, Y. Huang, L. Wei, F. Zhang, and H. Li, “Deep convolutional neural networks for hyperspectral image classification,” *J. Sensors*, vol. 2015, pp. 1–12, Jul. 2015.
- [11] H. Sun, X. Zheng, X. Lu, and S. Wu, “Spectral–spatial attention network for hyperspectral image classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3232–3245, May 2020.
- [12] L. Fang, G. Liu, S. Li, P. Ghamisi, and J. A. Benediktsson, “Hyperspectral image classification with squeeze multibias network,” *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 3, pp. 1291–1301, Mar. 2019.
- [13] Z. Zhong, J. Li, Z. Luo, and M. Chapman, “Spectral–spatial residual network for hyperspectral image classification: A 3D deep learning framework,” *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 847–858, Feb. 2018.
- [14] L. He, J. Li, C. Liu, and S. Li, “Recent advances on spectral–spatial hyperspectral image classification: An overview and new guidelines,” *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1579–1597, Mar. 2018.
- [15] F. Luo, L. Zhang, X. Zhou, T. Guo, Y. Cheng, and T. Yin, “Sparse-adaptive hypergraph discriminant analysis for hyperspectral image classification,” *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 6, pp. 1082–1086, Jun. 2020.
- [16] C. Yu *et al.*, “Hyperspectral image classification method based on CNN architecture embedding with hashing semantic feature,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 6, pp. 1866–1881, Jun. 2019.
- [17] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson, “Deep learning for hyperspectral image classification: An overview,” *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6690–6709, Sep. 2019.
- [18] W. Song, S. Li, L. Fang, and T. Lu, “Hyperspectral image classification with deep feature fusion network,” *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3173–3184, Jun. 2018.
- [19] N. Audebert, B. Le Saux, and S. Lefevre, “Deep learning for classification of hyperspectral data: A comparative review,” *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 159–173, Jun. 2019.
- [20] B. Liu, X. Yu, P. Zhang, A. Yu, Q. Fu, and X. Wei, “Supervised deep feature extraction for hyperspectral image classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 1909–1921, Apr. 2018.
- [21] G. Cheng, P. Zhou, and J. Han, “Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7405–7415, Dec. 2016.

- [22] J. Han, D. Zhang, G. Cheng, L. Guo, and J. Ren, "Object detection in optical remote sensing images based on weakly supervised learning and high-level feature learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3325–3337, Jun. 2015.
- [23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [24] L. Dong, Y. Yuan, and X. Lu, "Spectral-spatial joint sparse NMF for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, early access, Jul. 15, 2020, doi: [10.1109/TGRS.2020.3006109](https://doi.org/10.1109/TGRS.2020.3006109).
- [25] T. Ince, "Superpixel based graph Laplacian regularization for sparse hyperspectral unmixing," 2020, *arXiv:2007.14033*. [Online]. Available: <http://arxiv.org/abs/2007.14033>
- [26] P. Massoudifar, A. Rangarajan, and P. Gader, "Superpixel estimation for hyperspectral imagery," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2014, pp. 287–292.
- [27] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Semi-supervised hyperspectral image segmentation," in *Proc. 1st Workshop Hyperspectral Image Signal Process., Evol. Remote Sens.*, Aug. 2009, pp. 1–4.
- [28] Y. Chen, L. Zhu, P. Ghamisi, X. Jia, G. Li, and L. Tang, "Hyperspectral images classification with Gabor filtering and convolutional neural network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2355–2359, Dec. 2017.
- [29] L. Mou, P. Ghamisi, and X. X. Zhu, "Deep recurrent neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3639–3655, Jul. 2017.
- [30] H. Wu and S. Prasad, "Semi-supervised deep learning using pseudo labels for hyperspectral image classification," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1259–1270, Mar. 2018.
- [31] S. Hao, W. Wang, Y. Ye, T. Nie, and L. Bruzzone, "Two-stream deep architecture for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 2349–2361, Apr. 2018.
- [32] E. Aptoula, M. C. Ozdemir, and B. Yanikoglu, "Deep learning with attribute profiles for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 12, pp. 1970–1974, Dec. 2016.
- [33] L. Mou, P. Ghamisi, and X. Xiang Zhu, "Unsupervised spectral-spatial feature learning via deep residual Conv–Deconv network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 1, pp. 391–406, Jan. 2018.
- [34] F. Luo, H. Huang, Y. Duan, J. Liu, and Y. Liao, "Local geometric structure feature for dimensionality reduction of hyperspectral imagery," *Remote Sens.*, vol. 9, no. 8, p. 790, Aug. 2017.
- [35] X. Lu, H. Sun, and X. Zheng, "A feature aggregation convolutional neural network for remote sensing scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 7894–7906, Oct. 2019.
- [36] H. Sun, S. Li, X. Zheng, and X. Lu, "Remote sensing scene classification by gated bidirectional network," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 82–96, Jan. 2020.
- [37] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6232–6251, Oct. 2016.
- [38] J. Han, X. Yao, G. Cheng, X. Feng, and D. Xu, "P-CNN: Part-based convolutional neural networks for fine-grained visual categorization," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Aug. 6, 2020, doi: [10.1109/TPAMI.2019.2933510](https://doi.org/10.1109/TPAMI.2019.2933510).
- [39] X. Yao, J. Han, G. Cheng, X. Qian, and L. Guo, "Semantic annotation of high-resolution satellite images via weakly supervised learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3660–3671, Jun. 2016.
- [40] X. Zheng, Y. Yuan, and X. Lu, "Hyperspectral image denoising by fusing the selected related bands," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 5, pp. 2596–2609, May 2019.
- [41] X. Cao, F. Zhou, L. Xu, D. Meng, Z. Xu, and J. Paisley, "Hyperspectral image classification with Markov random fields and a convolutional neural network," *IEEE Trans. Image Process.*, vol. 27, no. 5, pp. 2354–2367, May 2018.
- [42] Z. Niu, W. Liu, J. Zhao, and G. Jiang, "DeepLab-based spatial feature extraction for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 2, pp. 251–255, Feb. 2019.
- [43] L. Jiao, M. Liang, H. Chen, S. Yang, H. Liu, and X. Cao, "Deep fully convolutional network-based spatial distribution prediction for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5585–5599, Oct. 2017.
- [44] H. Lee and H. Kwon, "Going deeper with contextual CNN for hyperspectral image classification," *IEEE Trans. Image Process.*, vol. 26, no. 10, pp. 4843–4855, Oct. 2017.
- [45] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.
- [46] X. Kang, C. Li, S. Li, and H. Lin, "Classification of hyperspectral images by Gabor filtering based deep network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 4, pp. 1166–1178, Apr. 2018.
- [47] P. Gader, A. Zare, R. Close, J. Aitken, and G. Tuell, "MUUFL Gulfport hyperspectral and Lidar airborne data set," Univ. Florida, Gainesville, FL, USA, Tech. Rep. REP-2013-570, 2013.
- [48] X. Du and A. Zare, "Technical report: Scene label ground truth map for MUUFL Gulfport data set," Univ. Florida, Gainesville, FL, Tech. Rep. 20170417, 2017.
- [49] W. Zhu *et al.*, "Unsupervised classification in hyperspectral imagery with nonlocal total variation and primal-dual hybrid gradient algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2786–2798, May 2017.
- [50] J. Zhu, L. Fang, and P. Ghamisi, "Deformable convolutional neural networks for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 8, pp. 1254–1258, Aug. 2018.
- [51] L. Fang, N. He, S. Li, A. J. Plaza, and J. Plaza, "A new spatial-spectral feature extraction method for hyperspectral images using local covariance matrix representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3534–3546, Jun. 2018.
- [52] D. Hong, X. Wu, P. Ghamisi, J. Chanussot, N. Yokoya, and X. X. Zhu, "Invariant attribute profiles: A spatial-frequency joint feature extractor for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 3791–3808, Jun. 2020.

Hao Sun is currently pursuing the Ph.D. degree with the Key Laboratory of Spectral Imaging Technology CAS, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, Shaanxi, China, and the University of Chinese Academy of Sciences, Beijing, China. His current research interests include pattern recognition, computer vision, and machine learning.



Xiangtao Zheng (Member, IEEE) received the M.Sc. and Ph.D. degrees in signal and information processing from the University of Chinese Academy of Sciences, China, in 2014 and 2017, respectively. He is currently an Associate Professor with the Key Laboratory of Spectral Imaging Technology CAS, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, Shaanxi, China. His main research interests include computer vision and pattern recognition.



Xiaoqiang Lu (Senior Member, IEEE) is currently a Full Professor with the Key Laboratory of Spectral Imaging Technology CAS, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, Shaanxi, China. His current research interests include pattern recognition, machine learning, hyperspectral image analysis, cellular automata, and medical imaging.

